AD-A048 027 MATERIALS RESEARCH LABS ASCOT VALE (AUSTRALIA) F/G 6/1
THE ACCURACY OF CURVE FITTING PROCEDURES FOR THE DETERMINATION --ETC(U)
JUL 77 R M DAWSON, W M HARPER, H A KELLOCK

UNCLASSIFIED MRL-R-695 NL

| OF |
ADA048027

END
DATE FILMED
1-78
DDC

MRL-R-695

AR-000-665



# DEPARTMENT OF DEFENCE

## DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION

## MATERIALS RESEARCH LABORATORIES

MELBOURNE, VICTORIA

## REPORT

## MRL-R-695

# THE ACCURACY OF CURVE FITTING PROCEDURES FOR THE DETERMINATION OF THE RATE CONSTANT FOR REACTIVATION OF AN ACYLATED ESTERASE

R.M. Dawson, W.M. Harper and H.A. Kellock



Approved for Public Release





JULY, 1977

DESTROY THIS REPORT WHEN IT IS
NO LONGER NEEDED.
DO NOT RETURN IT TO THE ORIGINATOR.

DEPARTMENT OF DEFENCE

MATERIALS RESEARCH LABORATORIES

# REPORT

MRL-R-695

# THE ACCURACY OF CURVE FITTING PROCEDURES FOR THE DETERMINATION OF THE RATE CONSTANT FOR REACTIVATION OF AN ACYLATED ESTERASE

R.M. Dawson, W.M. Harper [redacted] H.A. Kellock

ABSTRACT

Curve fitting procedures were examined for determination of the three unknown parameters A, B and k in the equation $y = A - Be^{-kt}$. Six methods were evaluated: one based on the sum of differences, three based on the sum of squares, and two based on the sum of squares for the equation in a linear form. It was concluded that no method was satisfactory. Even when an apparently accurate estimate of the rate constant k was obtained, it was accompanied by wide confidence limits. If possible, the constant A should be determined independently; satisfactory results can then be obtained.

Approved for Public Release

Jul 77

23 p.



POSTAL ADDRESS: Chief Superintendent, Materials Research Laboratories
P.O. Box 50, Ascot Vale, Victoria 3032, Australia

409014

## DOCUMENT CONTROL DATA SHEET

*Security classification of this page:* UNCLASSIFIED

1. *DOCUMENT NUMBERS:*
   a. *AR Number:* AR-000-665
   b. *Series & Number:* REPORT MRL-R-695
   c. *Report Number:* MRL-R-695

2. *SECURITY CLASSIFICATION:*
   a. *Complete document:* UNCLASSIFIED
   b. *Title in isolation:* UNCLASSIFIED
   c. *Abstract in isolation:* UNCLASSIFIED

3. *TITLE:*
   THE ACCURACY OF CURVE FITTING PROCEDURES FOR THE DETERMINATION OF THE RATE CONSTANT FOR REACTIVATION OF AN ACYLATED ESTERASE

4. *PERSONAL AUTHOR(S):*
   DAWSON, R.M.,
   HARPER, W.M. and
   KELLOCK, H.A.

5. *DOCUMENT DATE:* JULY, 1977

6. *TYPE OF REPORT & PERIOD COVERED:*

7. *CORPORATE AUTHOR(S):*
   Materials Research Laboratories

8. *REFERENCE NUMBERS:*
   a. *Task:*
   b. *Sponsoring Agency:*

9. *COST CODE:* 321250

10. *IMPRINT (Publishing establishment)*
    Materials Research Laboratories,
    P.O. Box 50, Ascot Vale, Vic.3032
    JULY, 1977

11. *COMPUTER PROGRAMME(S):*
    *(Title(s) and language(s)):*

12. *RELEASE LIMITATIONS (of the document):*

Approved for Public Release

12-0. *OVERSEAS:* N.O. [ ] P.R. [1] A [ ] B [ ] C [ ] D [ ] E [ ]

13. *ANNOUNCEMENT LIMITATIONS (of the information on this page):*

No Limitation

14. *DESCRIPTORS:*
    Curve fitting; Computer programs; Activation; Neostigmine; Acetylcholinesterase.

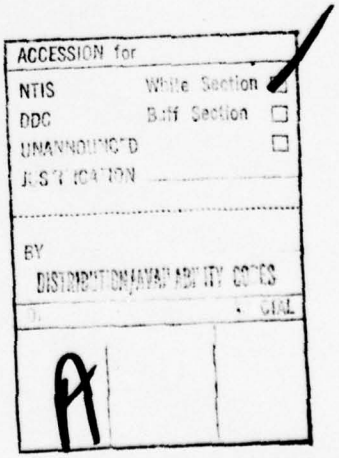
15. *COSATI CODES:* 1201, 0902, 0615

16. *ABSTRACT (if this is security classified, the announcement of this report will be similarly classified):*

Curve fitting procedures were examined for determination of the three unknown parameters A, B and k in the equation $y = A-Be^{-kt}$. Six methods were evaluated: one based on the sum of differences, three based on the sum of squares, and two based on the sum of squares for the equation in a linear form. It was concluded that no method was satisfactory. Even when an apparently accurate estimate of the rate constant k was obtained, it was accompanied by wide confidence limits. If possible, the constant A should be determined independently; satisfactory results can then be obtained.

# CONTENTS


| | Page No. |
|---|---|
| INTRODUCTION | 1 |
| METHODS | 2 |
| RESULTS AND DISCUSSION | 3 |
| CONCLUSION | 5 |
| REFERENCES | 6 |
| APPENDIX | 13 |

# THE ACCURACY OF CURVE FITTING PROCEDURES FOR THE DETERMINATION OF THE RATE CONSTANT FOR REACTIVATION OF AN ACYLATED ESTERASE

## INTRODUCTION

The solution of equations encountered in enzyme kinetics often requires the use of curve fitting procedures [1]. Recent work has emphasised that rigorous testing of such procedures is necessary before they can be utilised confidently in an experimental situation [2-4]. One example of an application in which curve fitting procedures fail to give reliable results is given in the present paper. The case under consideration is that of an esterase which has been inactivated by an acylating inhibitor. On removal of free inhibitor the enzyme recovers activity according to equation (1).

$$y = A-Be^{-kt} \qquad (1)$$

where y is the enzymic activity at time t and A, B and k are constants [5]. Specifically, A is the activity of completely re-activated (uninhibited) enzyme and (A-B) is the activity at t = 0. If A is known, equation (1) can be reduced to a linear form (equation 2) and the rate constant k and the constant B determined by least-squares linear regression on ln(A-y) vs t.

$$\ln(A-y) = \ln B-kt \qquad (2)$$

Often however the value of A is not known accurately and k can then be determined only by more complex curve fitting procedures [6]. A summary of the procedures likely to be adopted, and their accuracy, is given below. Simulated and experimental data were examined. The experiments showed that if all three parameters were determined by various curve fitting methods, very large errors could occur. Even in cases where apparently good results were obtained, error analyses showed that confidence ranges for k were very wide. If A was taken as a known constant, however, the fitted values of the other two parameters were much closer to their correct values and the confidence ranges for k were much narrower.

## METHODS

*Derivation of Simulated Data*

A set of random numbers between 0 and 1 was generated by a PDP 10 computer. Algebraic manipulation of the numbers produced two random error sets; one in which the numbers varied about zero (set E1) and the other in which the numbers varied about 1 (set E2). A third random number set (E3) was derived from a different set of computer-generated random numbers between 0 and 1. The three error sets are listed in Table I. Appropriate values were assigned to the constants A, B and k in equation (1), and 8 theoretical values of y thereby obtained (Table II). The number of theoretical values (8) was chosen as being a likely number of observations in an experimental situation. The simulated experimental values of y were then obtained from the theoretical values using one of the error sets in Table I. These simulated values and their method of calculation are also given in Table II. Several sets of simulated data were obtained in this manner (Table II).

*Experimental Data*

Bovine erythrocyte acetylcholinesterase (acetylcholine hydrolase, EC 3.1.1.7) was inhibited with neostigmine bromide (a phenolic ester of dimethylcarbamic acid). The solution of dimethylcarbamylated enzyme was diluted extensively to reduce the concentration of neostigmine to a non-inhibitory level and the enzymic activity was measured at various times after dilution by the method of Ellman et al. [7]. Full experimental details will be published later (R.M. Dawson, manuscript in preparation). In the presence of 1.8 mM choline chloride, the half-life for recovery of enzymic activity was observed to be approx. 13 min, and the activity at 180 min (14 half-lives) was taken to represent 100% enzymic activity (constant A in equation 1). The experimental data (Table III) were plotted in accordance with equation (2) and the constants B and k thereby determined by least-squares linear regression. Calculated values of y were then determined for each value of t under study using equation (1) and the "known" values of A, B and k, and the calculated values compared with the experimental values (Table III).

*Procedures for Curve Fitting*

All calculations were done on the computer using programs written in BASIC or FORTRAN. Six methods of curve fitting were investigated in order to find best-fit values of A, B and k. They were :

Method 1 : Minimise the standard error of the least-squares straight line through the data points $(\ln(A-y),t)$ [8].

Method 2 : Minimise an approximation to the standard error.

Method 3 : Minimise the sum of squares $\Sigma(y-\hat{y})^2$ [1].

Method 4 : Minimise $\Sigma[(y-\hat{y})^2/\hat{y}]$.

Method 5 : Minimise $\Sigma[(y-\hat{y})^2/(y+\hat{y})^2]$.

Method 6 : Minimise $\Sigma|y-\hat{y}|$.

In the above, y are the "experimental" values at known times t (Tables II and III) and $\hat{y}$ the values calculated from equation (1). Two methods of fitting by Method 1 were attempted, one by a linearisation technique (1A) [9] and the other by direct searching (1B).

In Method 3, it is assumed that the variance of y is independent of y [1] although there is no evidence that this is so for the experimental data set. For this reason, the sum of squares was weighted as shown in Methods 4 and 5; by $1/\hat{y}$ [10] or by $1/(y+\hat{y})^2$. The latter weighting factor, which was used in another situation, was claimed nevertheless to be free from bias and applicable to a variety of enzymic kinetic data [11].

Further details of the fitting procedures can be found in the Appendix.

## RESULTS AND DISCUSSION

*Description of Data Sets*

Before considering the results of the curve fitting procedures, it is worthwhile to summarise the characteristics of each data set, as follows. Sets I-III are simulated data sets (Table II).

Set I. A = 20, B = 19, k = 0.05. This set covers values of y from 0.26A to 0.87A. In set Ia, the sum of arithmetical errors (but not the sum of percentage errors) is almost zero, while in set Ib this situation is reversed.

Set II. A = 20, B = 19, k = 0.05. The values of the three constants are the same as in set I but the range covered by the values of y (0.18A to 0.71A) is not as great. Sets IIa and IIb bear the same relationship to each other, with respect to the distribution of simulated error, as sets Ia and Ib above. Sets I and II are not independent with respect to distribution of errors.

Set III. A = 20, B = 15, k = 0.02. This set covers a narrower range still of y values (0.32A to 0.66A). In set IIIa, neither the sum of arithmetical errors nor the sum of percentage errors approximates zero, and both are negative. In fact, 5 of the 8 data points exhibit a negative error, while in sets Ia, Ib, IIa, IIb and IV, the distribution of error is 4 positive, 4 negative.

Set IV. This is the experimental set and is listed in Table III. A = 27.8, B = 21.8, k = 0.053. The values of y cover the range 0.33A to 0.73A. Neither the sum of arithmetical errors nor the sum of percentage errors approximates zero, and both are negative. Set IV is therefore very similar to set IIIa in this respect.

The limited number of data sets above suffice for the purpose of this paper, which is to demonstrate the inaccuracies of the curve fitting procedures under study. Were some or all of these procedures acceptable, many more data sets would naturally be required before one procedure could be recommended in preference to others.

*Results from Curve Fitting*

The curve fitting procedures were applied to each of the 6 data sets under study. The best-fit values of the constants A, B and k for each of the six methods are given in Table IV where they are compared with the "true" values.

Both variations of Method 1 failed to give reasonable results and those which were obtained were inconsistent. It should be noted that the standard error S can be reduced to zero in this method by putting $A = B = \infty$ and $k = 0$. Local minima in S might be expected, but this occurred in only three of the six data sets using direct searching (Method IB). Furthermore, the variation of the standard error with the constant A in the vicinity of a local minimum (or maximum) was erratic. Data set IIIa actually gave a local *maximum* of S for a value of A (19.7) close to its "true" value (20.0). By Method IA, four of the six data sets gave convergence to the absolute minimum of $S = 0$, while the other two gave convergence to spurious values as shown in Table IV, probably because the rate of change in the corrected values of the parameters was slow enough to trigger off the program test of convergence. The failure of Method 1 is specially important because such an approach has been advocated as part of the determination of the kinetic constants of a mixture of chemically-modified and unmodified enzyme [8]. It should be noted however that if A is known, the objection to this method no longer applies as it then becomes a simple linear fit using equation (2). Results using this approach are shown in Table VI (see later).

Methods 2-5, unlike Method 1, always gave convergence. Superficially, Methods 3-5 were satisfactory for data sets Ia, Ib, IIa and IIb in that the per cent deviations of the best-fit values of A and B from their respective true values were less than the per cent deviation in some values of y, or nearly so. This is not surprising since there is a uniformity about the distribution of error in data sets I and II, and the values of y span a relatively wide range. Nevertheless the results are less satisfactory when the confidence ranges for the rate constant k are taken into consideration (see below and Table V).

The results are certainly unsatisfactory in cases IIIa and IV. In the former case, Methods 2-5 all gave essentially the same best-fit value of A (16.9 - 17.05) but this was lower than its true value by about 15%. Similarly for B and k the best-fit values differed from their true values by 18% and 43% respectively even though no "experimental" value of y differed from its true value by more than 2.33% (a very acceptable level of accuracy in many fields of science). In the case of data set IV which is a real experimental situation, the maximum "error" in y was 3.22%, yet the *best* values of A, B and k obtained by any method were in error by 6.3%, 4.9% and 19% respectively.

Method 6 was used as a trial only on sets Ia and IIIa but was not extended to other data sets because of unpromising results and the amount of computing required.

As a further test of the reliability of results obtained from our test data, the standard deviation of the fitted value of k was computed for Methods 2-5 using the calculation procedure described by Guest [9]. The values obtained for each method were all much the same. Table V shows the values obtained by Method 3 together with a 95% confidence range for k obtained by assuming that k is normally distributed. The results show how wide the confidence ranges are even in the case of superficially accurate values of k. Where apparently poor values were obtained for k, the true value was well inside the 95% confidence range. It should be noted that the results in Table V were obtained on the assumption of equal variances in the y values so that the sums of squares of residuals could be used in the estimations. The Appendix indicates how other assumptions can be used. Because of the sensitivity of the value of k to experimental error, the results reported by Hovanec & Lieske [6] (in which agreement of the computed constants A and k with those observed from test data was claimed to be within 2%) should not be taken as representative of the accuracy likely to be obtained in applications of this sort. They used a published computer program [12] which uses our Method 3 with a different calculation procedure for the analysis of their data using equation (1).

In order to show the advantage of knowing the value of A, the best-fit values of B and k together with standard deviations and 95% confidence ranges for k were determined for the six data sets assuming that the "true" values of A were known and linear fits using equation (2) could be made. The results are shown in Table VI. They indicate the better values obtained for B and k in this situation and the greater reliability to be placed on the results as shown by the tighter confidence ranges for k.

## CONCLUSION

Our results suggest that if an accurate determination of the rate constant in equation (1) is required, it is preferable to concentrate on determining the constant A experimentally, if necessary by an indirect method. An alternative, which would probably be more time-consuming, is to make a thorough investigation of the nature of the experimental error in the system under study [13]; however knowing this, the results of curve fitting procedures may still not be sufficiently accurate. This conclusion is important when it is realised that equation (1) is not confined to enzyme kinetics, but describes many processes in other fields of science [14] and engineering [12]. Curve fitting procedures have been adopted in these references [12,14] and also, in the field of enzyme reactivation, by Lieske et al. [15].

## REFERENCES


1. Cleland, W.W. (1967). Adv. Enzymol., 29, 1-32.

2. Newman, P.F.G., Atkins, G.L. and Nimmo, I.A. (1974). Biochem. J., 143, 779-781.

3. Atkins, G.L. and Nimmo, I.A. (1975). Biochem. J., 149, 775-777.

4. Nimmo, I.A. and Atkins, G.L. (1976). Biochem. J., 157, 489-492.

5. Aldridge, W.N. and Reiner, E. (1972). "Enzyme Inhibitors as Substrates", (Frontiers of Biology, vol. 26). North-Holland, Amsterdam.

6. Hovanec, J.W. and Lieske, C.N. (1972). Biochemistry, 11, 1051-1056.

7. Ellman, G.L., Courtney, K.D., Andres, V. and Featherstone, R.M. (1961). Biochem. Pharmacol., 7, 88-95.

8. Ashman, L.K. and Keech, D.B. (1975). Aust. J. Biol. Sci., 28, 379-387.

9. Guest, P.G. (1961). "Numerical Methods of Curve Fitting". Cambridge University Press.

10. Simeon, V. (1974). Croat. Chem. Acta, 46, 137-144.

11. Ottaway, J.H. (1973). Biochem. J., 134, 729-736.

12. Hayo, G.E. and Wilcoxon, W.L. (1963). "Exponential Curve Fitting with Applications". Technical Report R262, U.S. Naval Civil Engineering Laboratory, Port Hueneme, California.

13. Storer, A.C., Darlison, M.G. and Cornish-Bowden, A. (1975). Biochem. J., 151, 361-367.

14. Fisher, R.B. and Gardner, M.L.G. (1974). J. Physiol. (London), 241, 211-234.

15. Lieske, C.N., Clark, J.H., Lowe, J.R., Horton, G.L., Jewell, D.K. and Daasch, L.W. (1976). Edgewood Arsenal Technical Report EB-TR-76113. (U.S. Department of the Army, Aberdeen Proving Ground, Maryland 21010).

## TABLE I

### RANDOM ERROR SETS

| i | $(E1)_i$ | $(E2)_i$ | $(E3)_i$ |
|---|---|---|---|
| 1 | -0.036 | 0.993 | -0.141 |
| 2 | 0.307 | 1.061 | 0.098 |
| 3 | 0.027 | 1.005 | -0.101 |
| 4 | 0.095 | 1.019 | 0.232 |
| 5 | -0.229 | 0.954 | -0.018 |
| 6 | -0.108 | 0.978 | 0.134 |
| 7 | -0.166 | 0.967 | -0.159 |
| 8 | 0.105 | 1.021 | -0.170 |

## TABLE II

THEORETICAL ($y_{th}$) AND SIMULATED VALUES OF y FOR EQUATION (1)

t (time) = 5i for data sets I, Ia, Ib, III and IIIa.
t = 3i for data sets II, IIa, IIb.
For data sets Ia, II, A = 20, B = 19, k = 0.05
For data set III, A = 20, B = 15, k = 0.02

| Data Set | Nature of error | i | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| I | $y=y_{th}$ | 5.205 | 8.475 | 11.024 | 13.010 | 14.557 | 15.761 | 16.698 | 17.429 |
| Ia | $y=y_{th} + (E1)_i$ | 5.169 | 8.782 | 11.051 | 13.105 | 14.328 | 15.653 | 16.532 | 17.534 |
| Ib | $y=(y_{th}).(E2)_i$ | 5.168 | 8.996 | 11.083 | 13.258 | 13.891 | 15.420 | 16.143 | 17.794 |
| II | $y=y_{th}$ | 3.647 | 5.925 | 7.884 | 9.573 | 11.024 | 12.276 | 13.352 | 14.277 |
| IIa | $y=y_{th} + (E1)_i$ | 3.611 | 6.232 | 7.911 | 9.668 | 10.795 | 12.168 | 13.186 | 14.382 |
| IIb | $y=(y_{th}).(E2)_i$ | 3.621 | 6.289 | 7.926 | 9.755 | 10.520 | 12.010 | 12.908 | 14.576 |
| III | $y=y_{th}$ | 6.427 | 7.719 | 8.888 | 9.945 | 10.902 | 11.768 | 12.551 | 13.260 |
| IIIa | $y=y_{th} + (E3)_i$ | 6.286 | 7.817 | 8.786 | 10.177 | 10.884 | 11.902 | 12.392 | 13.090 |

## TABLE III

### EXPERIMENTAL DATA CONSISTENT WITH EQUATION (1)

$y_{th}$ = theoretical value of y assuming A = 27.8 (observed value) and B = 21.8383, k = 0.05315 (calculated from the data by least-squares linear regression, assuming A = 27.8; see equation 2). The units of y are relative ones; a value of y = 1 corresponds to an absorbance change at 412 nm of $1.33 \times 10^{-2}$ per min.

y = observed value of y

$100(y-y_{th})/y$ = percentage error

| Variable | t(min) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 3 | 5 | 7 | 10 | 13 | 15 | 17.5 | 20 |
| $y_{th}$ | 9.180 | 11.057 | 12.744 | 14.966 | 16.858 | 17.960 | 19.184 | 20.256 |
| y | 9.0 | 10.8 | 12.6 | 15.0 | 17.4 | 18.1 | 19.4 | 19.8 |
| $y-y_{th}$ | -0.180 | -0.257 | -0.144 | 0.034 | 0.542 | 0.140 | 0.216 | -0.456 |
| $100(y-y_{th})/y$ | -1.96 | -2.32 | -1.13 | 0.23 | 3.22 | 0.78 | 1.13 | -2.25 |

## TABLE IV

BEST-FIT VALUES OF THE CONSTANTS A, B AND k BY VARIOUS METHODS DESCRIBED IN THE TEXT

No figure given for Method 1B indicates that the standard error of the straight line in the logarithmic form decreased continually as A increased.

No figure given for Method 6 means that the method was not attempted.

The data sets (I-IV) are given in Tables II and III.

| Data Set | Parameter | True | Method 1A | Method 1B | Method 2 | Method 3 | Method 4 | Method 5 | Method 6 |
|---|---|---|---|---|---|---|---|---|---|
| Ia | A | 20.0 | 20.03 | 20.91 | 20.66 | 19.81 | 19.56 | 19.33 | 19.42 |
| | B | 19.0 | 18.78 | 18.91 | 18.86 | 18.68 | 18.64 | 18.56 | 18.59 |
| | 100k | 5.0 | 4.90 | 4.27 | 4.44 | 5.06 | 5.27 | 5.46 | 5.32 |
| Ib | A | 20.0 | ∞ | | 22.35 | 19.79 | 19.21 | 18.75 | |
| | B | 19.0 | ∞ | - | 19.50 | 18.34 | 18.23 | 18.09 | - |
| | 100k | 5.0 | o | | 3.48 | 4.92 | 5.41 | 5.83 | |
| IIa | A | 20.0 | 21.30 | 23.00 | 21.46 | 20.05 | 19.19 | 18.46 | |
| | B | 19.0 | 19.90 | 21.26 | 19.97 | 18.85 | 18.19 | 17.62 | - |
| | 100k | 5.0 | 4.35 | 3.74 | 4.27 | 4.89 | 5.35 | 5.79 | |
| IIb | A | 20.0 | ∞ | | 24.38 | 20.98 | 19.28 | 18.10 | |
| | B | 19.0 | ∞ | - | 22.47 | 19.52 | 18.17 | 17.23 | - |
| | 100k | 5.0 | o | | 3.35 | 4.40 | 5.21 | 5.92 | |
| IIIa | A | 20.0 | ∞ | 18.50 | 16.89 | 16.99 | 17.02 | 17.05 | 16.15 |
| | B | 15.0 | ∞ | 13.56 | 12.27 | 12.35 | 12.37 | 12.39 | 11.66 |
| | 100k | 2.0 | o | 2.32 | 2.92 | 2.87 | 2.86 | 2.85 | 3.34 |
| IV | A | 27.8 | ∞ | | 22.93 | 24.76 | 25.37 | 26.04 | |
| | B | 21.84 | ∞ | - | 19.12 | 19.84 | 20.23 | 20.73 | - |
| | 100k($min^{-1}$) | 5.31 | o | | 9.25 | 7.25 | 6.79 | 6.35 | |

## TABLE V

## STANDARD DEVIATIONS AND 95% CONFIDENCE RANGES FOR 100k USING METHOD 3 OF TEXT

| Data Set | 100k | | |
|---|---|---|---|
| | True | 95% C.R. | S.D. |
| Ia | 5.0 | 5.06 ± 0.96 | 0.37 |
| Ib | 5.0 | 4.92 ± 2.48 | 0.96 |
| IIa | 5.0 | 4.89 ± 1.73 | 0.67 |
| IIb | 5.0 | 4.40 ± 3.33 | 1.30 |
| IIIa | 2.0 | 2.87 ± 1.13 | 0.44 |
| IV | 5.32 | 7.25 ± 3.16 | 1.23 |

## TABLE VI

BEST-FIT VALUES OF B AND 100k FOR FIXED A, AND STANDARD DEVIATIONS AND 95% CONFIDENCE RANGES FOR 100k OBTAINED BY LINEAR FIT USING EQUATION (2)

| Data Set | Fixed A | B | | 100k | | |
|---|---|---|---|---|---|---|
| | | True | Best-fit | True | 95% C.R. | S.D. |
| Ia | 20.0 | 19.0 | 18.87 | 5.0 | 4.95 ± 0.26 | 0.11 |
| Ib | 20.0 | 19.0 | 18.93 | 5.0 | 4.92 ± 0.81 | 0.33 |
| IIa | 20.0 | 19.0 | 18.86 | 5.0 | 4.93 ± 0.24 | 0.10 |
| IIb | 20.0 | 19.0 | 18.87 | 5.0 | 4.90 ± 0.55 | 0.22 |
| IIIa | 20.0 | 15.0 | 14.88 | 2.0 | 1.95 ± 0.13 | 0.05 |
| IV | 27.8 | - | 21.84 | - | 5.31 ± 0.53 | 0.22 |

# APPENDIX

## 1. CURVE FITTING METHODS

The following are brief descriptions of the curve fitting methods used.

### Method 1

This method aims to minimise the standard error of the least-squares line through the n data points $[\ln(A-y_i), t_i]$. The minimum of S is sought by determining appropriate values of A, B and k where

$$S = \sum_{i=1}^{n} [\ln(A-y_i) - \ln B + kt_i]^2.$$

As stated in the text, S can be reduced to zero by putting k=0, B=A and then allowing A to increase indefinitely. As this is an unacceptable result for our purposes, a local minimum occurring elsewhere is sought. Two methods were used in an attempt to achieve this. The first method (Method 1A) is a linearisation technique [1]. This starts with approximate values $A_o$, $B_o$ and $k_o$ of the respective parameters and then tries to find small (assumed) corrections $\delta A$, $\delta B$ and $\delta k$. In this experiment, the initial values used were the "true" values of the parameters. Following Guest's procedure [1], the sum of squares to be minimised becomes

$$\sum_{i=1}^{n} \left(A_o - y_i\right)^2 \left[ y_i^1 + \frac{\delta A}{A_o - y_i} - \frac{\delta B}{B_o} + t_i \delta k \right]^2$$

where $\delta A$, $\delta B$, $\delta k$ are the quantities to be determined and

$$y_i^1 = \ln(A_o - y_i) - \ln B_o + k_o t_i .$$

Then improved values of A, B, k are determined by adding the corrections to the starting values. If more accuracy is required, the procedure can be iterated as often as desired using the improved values as new initial values.

In the other method (Method 1B) the minimum of S was sought by a direct search, in which values of A were the basis of search, corresponding values of B and k being obtained by the usual linear fitting technique. Because of the failure to achieve useful results, Methods 1A and 1B were abandoned as being of little use for this type of problem.

Method 2

Instead of minimising S as above, an approximation to S is obtained by starting as before with approximate values $A_o$, $B_o$, $k_o$ together with corrections $\delta A$, $\delta B$, $\delta k$. Then

$$S = \sum_{i=1}^{n} \left[\ln(A_o+\delta A-\hat{y}_i-\delta y_i)-\ln(B_o+\delta B)+(k_o+\delta k)t_i\right]^2$$

where $$\ln(A_o-\hat{y}_i) = \ln B_o - k_o t_i$$

and $$y_i = \hat{y}_i+\delta y_i \ .$$

Then, using first order approximations to the logarithms, there results

$$S \simeq \sum_{i=1}^{n} \left(\frac{\delta A-\delta y_i}{A_o-y_i} - \frac{\delta B}{B_o} + t_i \delta k\right)^2$$

The quantities $\delta A$, $\delta B$, $\delta k$ can now be found to minimise this approximation to S. The problem is linear but gives results inferior to those obtained from methods based on the exponential form.

Method 3

This method aims to find A, B, k to minimise S where

$$S = \sum_{i=1}^{n} \left(y_i - A + Be^{-kt_i}\right)^2$$

This problem is discussed fully by Guest [1] who uses the linearisation process outlined under Method 1A. Overall the results using this method were the best obtained. It gave the best results for data sets Ia,b and IIa. It was not far from best for data sets IIb and IIIa.

Method 4

Similar to Method 3 but using weights $1/\hat{y}_i$ where

$$\hat{y}_i = A - Be^{-kt_i}$$

Not as good overall as Method 3 for the data sets used.

METHOD 5

Similar to Method 3 but using weights $1/(y_i+\hat{y}_i)^2$. Not as good overall as Methods 3 and 4.

METHOD 6

This method finds A, B, k to minimise S where

$$S = \sum_{i=1}^{n} \left| y_i - A + Be^{-kt_i} \right|$$

Because of the awkward nature of S, direct search methods seem to be necessary. The amount of computing is large but can be reduced by the following procedure.

Consider B and k as fixed quantities and calculate

$$A_i = y_i + Be^{-kt_i}, \quad i = 1,2, ---, n .$$

Now rearrange the values of $A_i$ as $C_j$ where

$$C_1 < C_2 < \text{------} < C_n$$

Then for the particular values of B and k, S is minimum for A equal to the median of the $C_j$. The supreme minimum is then found by searching on B and k. The method was used only on data sets Ia and IIIa, but as results were not worth the amount of computing involved, the method was not tried with the other data sets.

2. ERROR ESTIMATES

The usual method of obtaining an estimate of variance in least-squares problems involves computing the residual sum of squares using the fitted parameter variables. This procedure is described by Guest and is applicable to all of Methods 1-5 described above. The variances of the parameter estimates can then be determined and hence confidence ranges can be calculated if assumptions about the statistical distributions of the estimates are made, e.g. that they are normally distributed. This procedure however involves the assumption that the variance (and if applicable the forms of distribution) of the y values is the same throughout. It is quite possible for this not to be so. In this case the following procedure may be useful. It refers to Method 3 but can easily be adapted for other methods.

The normal equations have the solution

$$\delta A = (QX-PY+NZ)/D,$$

$$\delta B = (PX-MY+LZ)/D,$$

$$\delta k = (NX-LY+KZ)/B_o D$$

where $D = nQ-SP+TN,$

$$K = nU-S^2,\ L=nV-ST,\ M = nW-T^2,$$

$$N = SV-TU,\ P = SW-TV,\ Q = UW-V^2$$

and $$S = \sum_{i=1}^{n} e^{-k_o t_i},\ T = \sum_{i=1}^{n} t_i e^{-k_o t_i},\ U = \sum_{i=1}^{n} e^{-2k_o t_i},$$

$$V = \sum_{i=1}^{n} t_i e^{-2k_o t_i},\ W = \sum_{i=1}^{n} t_i^2 e^{-2k_o t_i},$$

$$X = \sum_{i=1}^{n} \delta y_i,\ Y = \sum_{i=1}^{n} e^{-k_o t_i} \delta y_i,\ Z = \sum_{i=1}^{n} t_i e^{-k_o t_i} \delta y_i,$$

where $A_o$, $B_o$, $k_o$ are the initial estimates of A, B, k respectively and $\delta A$, $\delta B$, $\delta k$ are the corrections to be added. The $\delta y_i$ are defined by

$$\delta y_i = y_i - A_o + B_o e^{-k_o t_i}.$$

The solution can be expressed in the form

$$\delta A = \sum_{i=1}^{n} A_i \delta y_i,\ \delta B = \sum_{i=1}^{n} B_i \delta y_i,\ \delta k = \sum_{i=1}^{n} k_i \delta y_i \qquad (1)$$

where the coefficients in the summations are given by

$$DA_i = Q-(P-Nt_i)e^{-k_o t_i},$$

$$DB_i = P-(M-Lt_i)e^{-k_o t_i},$$

$$B_o Dk_i = N-(L-Kt_i)e^{-k_o t_i}.$$

Any information or assumptions about the variability of the $y_i$, which is the same as that of the $\delta y_i$, can be used in the usual ways to obtain corresponding variabilities of the parameter estimates. For example, if the variances of the $y_i$ are $\sigma_i^2$, the variance $\sigma_k^2$ of the estimate of k will be estimated in the usual way by

$$\sigma_k^2 = \sum_{i=1}^{n} k_i^2 \sigma_i^2$$

and similarly for the other parameters. Systematic errors can also be dealt with if required.

As a numerical example, data set IIIa gives for the $\delta k$ expansion, using the "true" values of A, B, k for $A_o$, $B_o$, $k_o$

$$100\ \delta k = -1.721\ \delta y_1 + 0.015\ \delta y_2 + 0.983\ \delta y_3 + 1.312\ \delta y_4 + 1.116\ \delta y_5 + 0.492\ \delta y_6 - 0.477\ \delta y_7 - 1.720\ \delta y_8 \qquad (2)$$

and similar expressions for $\delta A$, $\delta B$.

From (2), results such as the following can be derived.

(a) Substituting the actual values of the $\delta y_i$ for this data set gives the value 0.00862 for $\delta k$, the increment to be added to $k_o$ on the first iteration of method 3, giving k = 0.0286.

(b) If the greatest possible error in the $y_i$ is ± 0.25 and the errors are cumulative, then

$$\left|\delta k\right| \leqslant 0.25 \sum_{i=1}^{n} \left|k_i\right|$$

giving a possible range for k of 0.0286 ± 0.0196.

(c) If it is assumed that the standard deviations of the $y_i$ are the same and that the maximum error is ± 0.25 and that this represents the 95% confidence limits for the $y_i$ values, then the variance $\sigma_k^2$ of $\delta k$ and hence of the k estimate is given by

$$\sigma_k^2 = \left(\frac{0.25}{1.96}\right)^2 \sum_{i=1}^{n} k_i^2 = 0.0000168$$

so that the 95% confidence range for k is 0.0286 ± 0.0105.

(d) If it is assumed that the standard deviations of the $y_i$ are proportional to the true values of the $y_i$ and that the 95% confidence range for each $y_i$ is ± 2.5% of the true value, then

$$\sigma_k^2 = \left(\frac{0.025}{1.96}\right)^2 \sum k_i^2 \hat{y}_i^2 = 0.0000180$$

so that the 95% confidence range for k is 0.0286 ± 0.0109. In an experimental situation, the "true" values of the $y_i$ would be unknown, but in this case the experimental values could be used without changing the results significantly.

REFERENCE

1. Guest, P.G. (1961). "Numerical methods of Curve Fitting". Cambridge University Press.

(MRL-R-695)

DISTRIBUTION LIST

MATERIALS RESEARCH LABORATORIES

Chief Superintendent
Superintendent, Organic Chemistry Division
Superintendent, Physics Division
Dr. Shirley E. Freeman
Mr. S.A. Lott
Dr. R.M. Dawson
Mr. H.A. Kellock
Mr. W.M. Harper
Library
Librarian, N.S.W. Branch (Through Officer-in-Charge)
Officer-in-Charge, Joint Tropical Research Unit

DEPARTMENT OF DEFENCE

Chief Defence Scientist
Executive Controller, ADSS
Superintendent, Defence Science Administration, DSTO
Superintendent, Military Advisers Branch
Head, Laboratory Programs Branch
Army Scientific Adviser
Air Force Scientific Adviser
Naval Scientific Adviser
Chief Superintendent, Aeronautical Research Laboratories
Director, Weapons Research Establishment
Senior Librarian, Weapons Research Establishment
Librarian, R.A.N. Research Laboratory
Document Exchange Centre, DIR (16 copies)
Principal Librarian, Campbell Park Library ADSATIS Annex
Directorate of Quality Assurance (Air Office)
Head, Engineering Development Establishment
Director, Joint Intelligence Organisation

DEPARTMENT OF PRODUCTIVITY

NASA Canberra Office
Head, B.D.R.S.S. (Aust.)

OTHER FEDERAL AND STATE DEPARTMENTS AND INSTRUMENTALITIES

The Chief Librarian, Central Library, C.S.I.R.O.
Australian Atomic Energy Commission Research Establishment
Chief, Division of Applied Organic Chemistry, C.S.I.R.O.
Chief, Division of Chemical Technology, C.S.I.R.O.
Chief, Division of Chemical Physics, C.S.I.R.O.

DISTRIBUTION LIST

(Continued)

MISCELLANEOUS - AUSTRALIA

The Librarian, Monash University
The Librarian, La Trobe University
The Librarian, Barr Smith Library, University of Adelaide
Prof. Sir S. Sunderland, Department of Experimental Neurology
Sir Hugh Ennor, Secretary, Department of Science

MISCELLANEOUS - OVERSEAS

Defence Scientific & Technical Representative, Department of Defence, England
Assistant Director/Armour and Materials, Military Vehicles and Engineering Establishment, England
Reports Centre, Directorate of Materials Aviation, England
Library - Exchange Desk, National Bureau of Standards, U.S.A.
U.S. Army Standardization Group, Office of the Scientific Standardization Representative, Canberra, A.C.T.
Senior Standardization Representative, U.S. Army Standardization Group, Canberra, A.C.T.
Chief, Research and Development, Defence Scientific Information Service, Canada (2 copies)
The Director, Defence Scientific Information and Documentation Centre, India
Colonel B.C. Joshi, Military, Naval and Air Adviser, High Commission of India, Red Hill, A.C.T.
Director, Defence Research Centre, Malaysia
Accessions Department, British Library, England
Official Publications Section, British Library, England
Librarian, Periodicals Recording Section, National Reference Library of Science and Invention, England
INSPEC: Acquisition Section, Institution of Electrical Engineers, England
Overseas Reports Section, Defence Research Information Centre, England
Editor, "New Scientist", England
Editor, "Nature", England
Director, Chemical Defence Establishment, England.